# PRINCIPLES OF VALIDATION OF DIAGNOSTIC ASSAYS FOR INFECTIOUS DISEASES[1]

R. H. JACOBSON,
New York State College of Veterinary Medicine,
Cornell University, Ithaca,
New York, USA

## Abstract

PRINCIPLES OF VALIDATION OF DIAGNOSTIC ASSAYS FOR INFECTIOUS DISEASES.

Assay validation requires a series of inter-related processes. Assay validation is an *experimental process*: reagents and protocols are optimized by experimentation to detect the analyte with accuracy and precision. Assay validation is a *relative process*: its diagnostic sensitivity and diagnostic specificity are calculated relative to test results obtained from reference animal populations of known infection/exposure status. Assay validation is a *conditional process*: classification of animals in the target population as infected or uninfected is conditional upon how well the reference animal population used to validate the assay represents the target population; accurate predictions of the infection status of animals from test results (PV+ and PV–) are conditional upon the estimated prevalence of disease/infection in the target population. Assay validation is an *incremental* process: confidence in the validity of an assay increases over time when use confirms that it is robust as demonstrated by accurate and precise results; the assay may also achieve increasing levels of validity as it is upgraded and extended by adding reference populations of known infection status. Assay validation is a *continuous process*: the assay remains valid only insofar as it continues to provide accurate and precise results as proven through statistical verification. Therefore, the work required for validation of diagnostic assays for infectious diseases does not end with a time-limited series of experiments based on a few reference samples rather, to assure valid test results from an assay requires constant vigilance and maintenance of the assay, along with reassessment of its performance characteristics for each unique population of animals to which it is applied.

## INTRODUCTION

Validation is the evaluation of a process to determine its fitness for a particular use. A validated assay yields test results that identify the presence of a particular analyte (e.g an antibody) and allows predictions to be made about the status of the test subject. However, for infectious disease diagnostic assays, the identity and definition of the criteria required for assay validation are elusive, and the process leading to a validated assay is not standardised.

By considering the variables that affect an assay's performance, the criteria that must be addressed in assay validation become clearer. The variables can be grouped into three categories: (a) the sample — host/organism interactions affecting the analyte composition and concentration in the serum sample; (b) the assay system — physical, chemical, biological and technician-related factors affecting the capacity of the assay to detect a specific analyte in the sample; and (c) the test result — the capacity of a test result, derived from the assay system, to predict accurately the status of the host relative to the analyte in question.

Factors that affect the concentration and composition of analyte in the serum sample are mainly attributable to the host and are either inherent (e.g. age, sex, breed, nutritional status, pregnancy, immunological responsiveness) or acquired (e.g. passively acquired antibody, active immunity elicited by vaccination or infection). Non-host factors, such as contamination or deterioration of the sample, may also affect the analyte in the sample.

Factors that interfere with the analytical accuracy of the assay system are instrumentation and technician error, reagent choice and calibration, accuracy of controls, reaction vessels, water quality, pH and ionicity of buffers and diluents, incubation temperatures and durations, and error introduced by detection of closely related analytes, such as antibody to cross-reactive organisms, rheumatoid factor, or heterophile antibody.

Factors that influence the capacity of the test result to predict accurately the infection or analyte status of the host[2] are diagnostic sensitivity (D-SN), diagnostic specificity (D-SP), and

---

[1] This paper has been published in the "Manual of Standards for Diagnostic Tests and Vaccines", p. 8-15 by OIE, France 1996 and is included as an excerpt in this TECDOC with the permission of the OIE.

prevalence of the disease in the population targeted by the assay. D-SN and D-SP are derived from test results on samples obtained from selected reference animals. The degree to which the reference animals represent all of the host and environmental variables in the population targeted by the assay has a major impact on the accuracy of test result interpretation. For example, experienced diagnosticians are aware that an assay, validated by using samples from northern European cattle, may not give valid results for the distinctive populations of cattle in Africa.

The capacity of a positive or negative test result to predict accurately the infection status of the animal is the most important consideration of assay validation. This capacity is not only dependent on a highly precise and accurate assay and carefully derived estimates of D-SN and D-SP, but is heavily influenced by prevalence of the infection in the targeted population. Without a current estimate of the disease prevalence in that population, the interpretation of a positive or negative test result may be compromised.

Many variables obviously must be addressed before an assay can be considered 'validated' [1]. However, there is no consensus whether the concept of assay validation is a time-limited process during which only those factors intrinsic to the assay are optimised and standardised, or whether the concept includes an ongoing validation of assay performance for as long as the assay is used. Accordingly, the term 'validated assay' elicits various interpretations among laboratory diagnosticians and veterinary clinicians. Therefore, a working definition of assay validation is offered as a context for the guidelines outlined below.

## A. DEFINITION OF ASSAY VALIDATION

A validated assay consistently provides test results that identify animals as positive or negative for an analyse or process (e.g. antibody, antigen, or induration at skin test site) and, by inference, accurately predicts the infection status of animals with a predetermined degree of statistical certainty. This chapter will focus on the principles underlying development and maintenance of a validated assay. Guidelines for the initial stages in assay development are included because they constitute part of the validation process. How this early process is carried out heavily influences the capacity of the eventual test result to provide diagnostic accuracy.

## B. STAGES OF ASSAY VALIDATION

Development and validation of an assay is an incremental process consisting of at least five stages: 1) Determination of the feasibility of the method for a particular use; 2) Choice, optimisation, and standardisation of reagents and protocols; 3) Determination of the assay's performance characteristics; 4) Continuous monitoring of assay performance; and 5) Maintenance and enhancement of validation criteria during routine use of the assay (Figure 1 ). Although some scientists may question the relevance of the fourth and fifth stages as validation criteria, they are included here because an assay can be considered valid only to the extent that test results are valid, i.e. whether they fall within statistically defined limits and provide accurate inferences.

An indirect enzyme-linked immunosorbent assay (ELISA) test for detection of antibody will be used in this chapter to illustrate the principles of assay validation. It is a test format that can be difficult to validate because of signal amplification of both specific and non-specific components [2]. This methodology serves to highlight the problems that need to be addressed in any assay validation process. The same basic principles are used in validation of other complex or simple assay formats.

Because of space limitations, this introductory chapter provides only basic guidelines for the principles concerned with assay validation. It is derived from a more detailed treatment of the subject [3].

### STAGE 1. FEASIBILITY STUDIES

In the ELISA example, feasibility studies are the first stage in validating a new assay. They are carried out in order to determine if the selected reagents and protocol have the capacity to

---

[2] In this paper, the terms positive and negative have been reserved for test result and never refer to infection or antibody/antigen status of the host. Whenever reference is made to 'infection' or 'analyte', any method of exposure to an infectious agent that could be detected directly (e.g. antigen) or indirectly (e.g. antibody) by an assay, should be inferred.

distinguish between a range of antibody concentrations to an infectious agent while providing minimal background activity. Feasibility studies also give initial estimates of repeatability, and of analytical sensitivity and specificity.

## 1. Control samples

It is useful to select four or five samples (serum in our example) that range from high to low levels of antibodies against the infectious agent in question. In addition, a sample containing no antibody is required. These samples will be used to optimise the assay reagents and protocol during feasibility studies, and later as control samples. The samples ideally should represent known infected and uninfected animals from the population that eventually will become the target of the validated assay. The samples should have given expected results in one or more serological assays other than the one being validated. The samples are preferably derived from individual animals, but they may represent pools of samples from several animals. A good practice is to prepare a large volume (e.g. 10 ml) of each sample and divide it into 0.1 ml aliquots for storage at -20$^{o}$C. One aliquot of each sample is thawed, used for experiments, and held at 4$^{o}$C between experiments until depleted. Then, another is thawed for further experimentation. This method provides the same source of serum with the same number of freeze-thaw cycles for all experiments (repeated freezing and thawing of serum can denature antibodies so should be avoided). Also, variation is reduced when the experimenter uses the same source of serum for all experiments rather than switching among various sera between experiments. This approach has the added advantage of generating a data trail for the repeatedly run samples. After the initial stages of assay validation are completed, one or more of the samples can become the serum control(s) that are the basis for data expression and repeatability assessments both within and between runs of the assay. They may also serve as standards if their activity has been predetermined; such standards provide assurance that runs of the assay are producing accurate data [1].

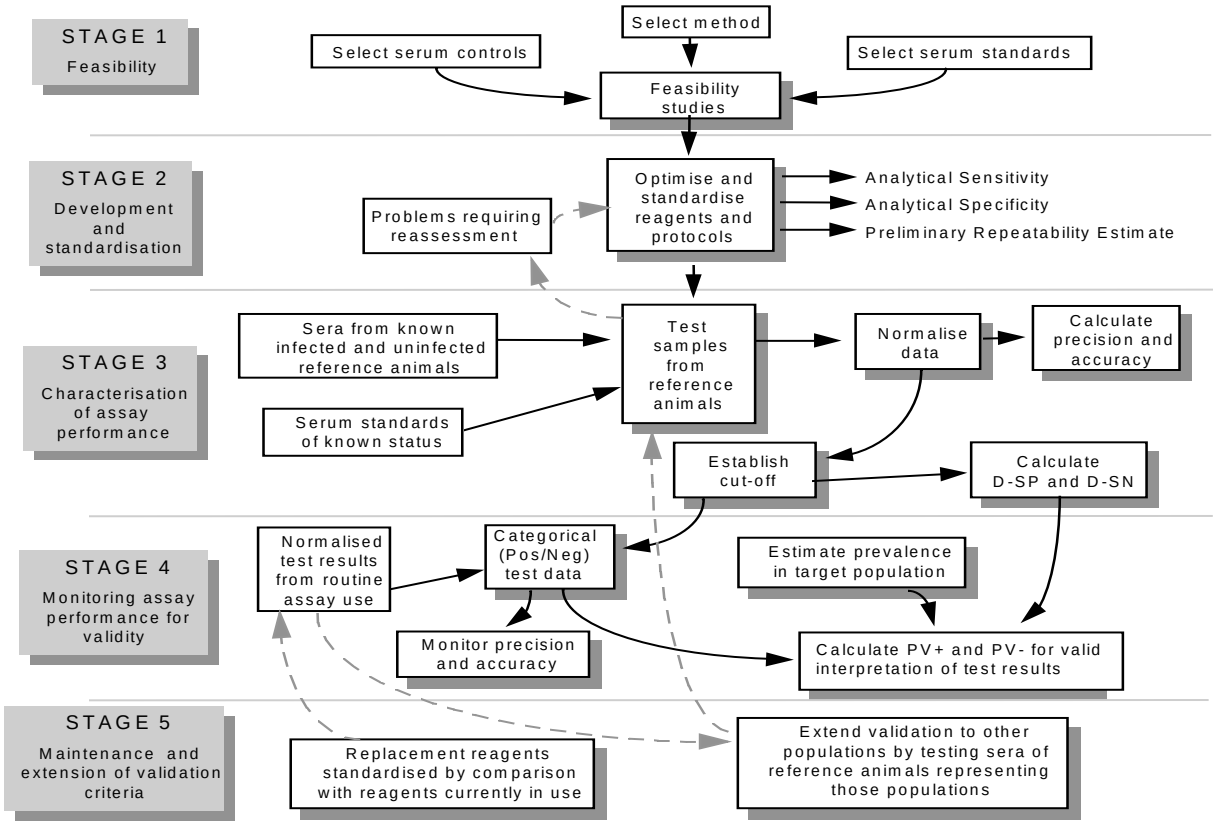


*FIG. 1. Stages in the incremental process of assay validation. Shadowed boxes indicate action points within each stage in the process.*

## 2. Selection of method to achieve normalised results

Normalisation adjusts raw test results of all samples relative to values of controls included in each run of the assay (not to be confused with transformation of data to achieve a 'normal' [Gausian] distribution). The method of normalisation and expression of data should be determined preferably no later than at the end of the feasibility studies. Comparisons of results from day to day and between laboratories are most accurate when normalised data are used. For example, in ELISA systems, raw optical density (absorbance) values are absolute measurements that are influenced by ambient temperatures, test parameters, and photometric instrumentation. To account for this variability, results are expressed as a function of the reactivity of one or more serum control samples that are included in each run of the assay. Data normalisation is accomplished in the indirect ELISA by expressing absorbance values in one of several ways [1]. A simple and useful method is to express all absorbance values as a percentage of a single high-positive serum control that is included on each plate. This method is adequate for most applications. More rigour can be brought to the normalisation procedure by calculating results from a standard curve generated by several serum controls. It requires a more sophisticated algorithm, such as linear regression or log-logit analysis. This approach is more precise because it does not rely on only one high-positive control sample for data normalisation, but rather uses several serum controls, adjusted to expected values, to plot a standard curve from which the sample value is extrapolated. This method also allows for exclusion of a control value that may fall outside expected confidence limits.

For assays that are end-pointed by sample titration, such as serum (viral) neutralisation, each run of the assay is accepted or rejected based on whether control values fall within predetermined limits. Because sample values usually are not adjusted to a control value, the data are not normalised by the strict definition of the term.

Whatever method is used for normalisation of the data, it is essential to include additional controls for any reagent that may introduce variability and thus undermine attempts to achieve a validated assay. The normalised values for those controls need to fall within predetermined limits (e.g. within ±2 or ±3 standard deviations of the mean of many runs of each control).

## STAGE 2. ASSAY DEVELOPMENT AND STANDARDISATION

When the feasibility of the method has been established the next step is to proceed with development of the assay and standardise the selected reagents and protocols.

### 1. Selection of optimal reagent concentrations and protocol parameters

Optimal concentrations/dilutions of the antigen adsorbed to the plate, serum, enzyme-antibody conjugate, and substrate solution are determined through 'checkerboard' titrations of each reagent against all other reagents, following confirmation of the best choice of reaction vessels (usually evaluation of two or three types of microtitre plates, each with its different binding characteristics, to minimise background activity while achieving the maximum spread in activity between negative and high-positive samples). Additional experiments determine the optimal temporal, chemical, and physical variables in the protocol, including incubation temperatures and durations; the type, pH, and molarity of diluent, washing, and blocking buffers; and equipment used in each step of the assay (for instance pipettes and washers that give the best reproducibility).

Optimisation of the reagents and protocol includes an assessment of accuracy by inclusion, in each run of the assay, one or more serum standards of a known level of activity for the analyte in question. An optimised assay that repeatedly achieves the same results for a serum standard and the serum controls may be designated as a standardised assay.

### 2. Repeatability - preliminary estimates

Preliminary evidence of repeatability (agreement between replicates within and between runs of the assay) is necessary to warrant further development of the assay. This is accomplished by evaluating results from replicates of all samples in each plate (intra-plate variation), and inter-plate variation using the same samples run in different plates within a run and between runs of the assay.

For ELISA, raw absorbance values are usually used at this stage of validation because it is uncertain whether the results of the high-positive control serum, which could be used for calculating normalised values, are reproducible in early runs of the assay format. Also, expected values for the controls have not yet been established. Three-to-four replicates of each sample run in at least five plates on five separate occasions are sufficient to provide preliminary estimates of repeatability. Coefficients of variation (standard deviation of replicates/mean of replicates), generally less than 20% for raw absorbance values, indicates adequate repeatability at this stage of assay development. However, if evidence of excessive variation (>30%) is apparent for most samples within and/or between runs of the assay, more preliminary studies should be done to determine whether stabilisation of the assay is possible, or whether the test format should be abandoned. This is important because an assay that is inherently variable has a high probability of not withstanding the rigours of day-to-day testing on samples from the targeted population of animals.

### 3. Determination of analytical sensitivity and specificity

The analytical sensitivity of the assay is the smallest detectable amount of the analyte in question, and analytical specificity is the degree to which the assay does not cross-react with other analytes. These parameters are distinguished from diagnostic sensitivity and specificity as defined below. Analytical sensitivity can be assessed by end-point dilution analysis, which indicates the dilution of serum in which antibody is no longer detectable. Analytical specificity is assessed by use of a panel of sera derived from animals that have experienced related infections that may stimulate cross-reactive antibodies. If, for instance, the assay does not detect antibody in limiting dilutions of serum with the same efficiency as other assays, or cross-reactivity is common when sera from animals with closely related infections are tested, the reagents need to be recalibrated or replaced, or the assay should be abandoned.

## STAGE 3. DETERMINING ASSAY PERFORMANCE CHARACTERISTICS

If the feasibility and initial development and standardisation studies indicate that the assay has potential for field application, the next step is to identify the assay's performance characteristics.

### 1. Diagnostic sensitivity and specificity

#### a) Principles and definitions

Estimates of diagnostic sensitivity (D-SN) and diagnostic specificity (D-SP) are the primary parameters obtained during validation of an assay. They are the basis for calculation of other parameters from which inferences are made about test results. Therefore, it is imperative that estimates of D-SN and D-SP are as accurate as possible. Ideally, they are derived from testing a series of samples from reference animals of known history and infection status relative to the disease/infection in question. Diagnostic sensitivity is the proportion of known infected reference animals that test positive in the assay; infected animals that test negative are considered to have false negative results. Diagnostic specificity is the proportion of uninfected reference animals that test negative in the assay; uninfected reference animals that test positive are considered to have false-positive results. The number and source of reference samples used to derive DSN and D-SP are of paramount importance if the assay is ever to be properly validated for use in the general population of animals targeted by the assay.

It is possible to calculate the number of reference samples, from animals of known infection/exposure status, required for determinations of D-SN and D-SP that will have statistically defined limits. Formulae for these calculations, their limitations, and a discussion of the selection criteria for standard sera are detailed elsewhere [3]. Because of the many variables that must be accounted for, at least 300 reference samples from known-infected animals, and no fewer than 1,000 samples from known-uninfected animals, should be included to determine initial estimates of D-SN and D-SP, respectively. The number of samples should be expanded to approximately 1,000 and 5,000, respectively, to establish precise estimates of D-SN and D-SP [3].

### b) Standards of comparison for the new assay

In serology, the 'standard of comparison' is the results of a method or combination of methods with which the new assay is compared. Although the term 'gold standard' is commonly used to describe any standard of comparison, it should be limited to methods that unequivocally classify animals as infected or uninfected. Some isolation methods themselves have problems of repeatability and sensitivity. Gold standard methods include unequivocal isolation of the agent or pathognomonic histopathological criteria. Because the gold standard is difficult to achieve, relative standards of comparison are often necessary; these include results from other serological assays and from experimentally infected or vaccinated animals. Calculations of D-SN and D-SP are most reliable when the gold standard of comparison is available. When only relative standards of comparison are available, estimates of D-SN and D-SP for the new assay may be compromised because the error in the estimates of D-SN and D-SP for the relative standard is carried over into those estimates for the new assay.

### c) Precision, repeatability, reproducibility, and accuracy

Repeatability and reproducibility are estimates of precision in the assay. Precision is a measure of dispersion of results for a repeatedly tested sample; a small amount of dispersion indicates a precise assay. Repeatability in a diagnostic assay has two elements: the amount of agreement between replicates (usually two or three) of each sample within a run of the assay, and the amount of between run agreement for the normalised values of each control sample. Reproducibility is the amount of agreement between results of samples tested in different laboratories. Accuracy is the amount of agreement between a test value and the expected value for an analyte in a standard sample of known activity (e.g., titre or concentration). An assay system may be precise, but not accurate, if the test results do not agree with the expected value for the standard.

Reliable estiates of repeatability and accuracy, both within and between runs of the assay, can be obtained by use of normalised results from the many runs of the new assay that were required to assess the sera of reference animals (less reliable estimates were obtained from preliminary data using raw absorbance values). At least 10, and preferably 20 runs of the assay will give reasonable initial estimates of these parameters. Methods for evaluating these parameters have been described in detail [3].

Accuracy can be assessed by inclusion of one or more standards (samples of known titre, concentration, etc.) in each run of the assay. The standards may be control sera provided that the amount of analyte (e.g. titre, concentration) in each one has been previously determined by comparison with primary or secondary reference standards [1], and the control sera are not used in the data normalisation process.

Reproducibility of the assay is determined in several laboratories using the identical assay (protocol, reagents, and controls) on a group of at least 10 samples, preferably duplicated to a total of 20 samples. These samples need to represent the full range of expected analyte concentrations in samples from the target population. The extent to which the collective results for each sample deviate from expected values is a measure of assay reproducibility. The degree of concordance of between laboratory data is one more basis for determining whether the assay's performance characteristics are adequate to constitute a validated assay.

## 2. Selection of the cut-off (positive/negative threshold)

To achieve estimates of D-SN and D-SP of the new assay, the test results first must be reduced to positive or negative categories. This is accomplished by insertion of a cut-off point (threshold or decision limit) on the continuous scale of test results. Although many methods have been described for this purpose, three examples will illustrate different approaches, together with their advantages and disadvantages. The first is a cut-off based on the frequency distributions [3] of test results from uninfected and infected reference animals. This cut-off can be established by visual inspection of the frequency distributions, by receiver-operator characteristics (ROC) analysis [4, 5], or by selection that favours either D-SN or D-SP, whichever is required for a given assay [6]. A second approach is establishing a cut-off based only on uninfected reference animals; this provides an

estimate of DSP but not D-SN. The third method provides an 'intrinsic cut-off' based on test results from sera drawn randomly from within the target population with. no prior knowledge of the animals' infection status [7]. Although no estimates of D-SN and D-SP are obtained by this method, they can be determined as confirmatory data are accumulated.

If considerable overlap occurs in the distributions of test values from known infected and uninfected animals, it is difficult to select a cut-off that will accurately classify these animals according to their infection status. Rather than a single cut-off, two cut-offs can be selected that define a high D-SN (e.g. inclusion of 99% of the values from infected animals), and a high D-SP (e.g. 99% of the values from uninfected animals). The values that fall between these percentiles would then be classified as suspicious or equivocal, and would require testing by a confirmatory assay or retesting for detection of seroconversion.

### 3. Calculation of diagnostic sensitivity and specificity

The selection of a cut-off allows classification of test results into positive or negative categories. Calculation of D-SN and D-SP are aided by associating the positive/negative categorical data with the known infection status for each animal using a two-way (2 x 2) table (Table 1). After the cut-off is established, results of tests on standard sera can be classified as true positive (TP) or true negative (TN) if they are in agreement with those of the gold standard (or other standard of comparison). Alternatively, they are classified as false positive (FP) or false negative (FN) if they disagree with the standard. Diagnostic sensitivity is calculated as TP/(TP + FN) whereas diagnostic specificity is f TN/(TN + FP); the results of both calculations are usually expressed as percentages (Table I).

TABLE I. CALCULATIONS OF D-SN AND D-SP AIDED BY A 2 X 2 TABLE THAT ASSOCIATES INFECTION STATUS WITH TEST RESULTS FROM 2000 REFERENCE ANIMALS

| | | **Reference animals of known infection status** | |
|---|---|---|---|
| | | **Infected** | **Uninfected** |
| **Test result** | **Positive** | **570** **TP** | **FP** **46** |
| | **Negative** | **30** **FN** | **TN** **1354** |
| | | **Diagnostic sensitivity** $\frac{TP}{TP+FN} = \frac{570}{600} = 95\%$ | **Diagnostic specificity** $\frac{TN}{TN+FP} = \frac{1354}{1400} = 96{,}7\%$ |

## STAGE 4. MONITORING VALIDITY OF ASSAY PERFORMANCE

### 1. Interpretation of test results - factors affecting assay validity

An assay's test results are useful only if the inferences made from them are accurate. A common error is to assume that an assay with 99% D-SN and 99% D-SP will generate one false-positive and one false-negative result for approximately every 100tests on animals from the target population. Such an assay may be precise and accurate, but produce test results that do not accurately predict infection status. For example, if the prevalence of disease in a population targeted by the assay

is only 1 per 1,000 animals, and the false positive test rate is 1 per 100 animals (99% D-SP), for every 1,000 tests on that population, ten will be false positive and one will be true positive. Hence, only approximately 9% of positive test results will accurately predict the infection status of the animal; the test result will be wrong 91% of the time. This illustrates that the capacity of a positive or negative test result to predict infection status is dependent on the prevalence of the infection in the target population [8]. Of course, the prevalence will probably have been determined by use of a serological test.

An estimate of prevalence in the target population is necessary for calculation of the predictive values of positive (PV+) or negative (PV-) test results. When test values are reported without providing estimates of the assay's D-SP and D-SN, it is not possible to make informed predictions of infection status from test results [8]. It is, therefore, highly desirable to provide an interpretation statement with test results accompanied by a small table indicating PV+ and PV- for a range of expected prevalences of infection in the target population. Without provision of such information, test results from the assay may have failed to accurately classify the infection status of animals, and thus do not reflect a fully validated assay.

### STAGE 5. MAINTENANCE AND ENHANCEMENT OF VALIDATION CRITERIA

A validated assay needs constant monitoring and maintenance to retain that designation. Once the assay is put into routine use, internal quality control is accomplished by consistently monitoring the assay for assessment of repeatability and accuracy [9].

Reproducibility between laboratories should be assessed at least twice each year. It is useful to volunteer membership in a consortium of laboratories that are interested in evaluating their output. In the near future, good laboratory practice, including implementation of a total quality assurance programme, will become essential for laboratories seeking to meet national and international certification requirements.

Proficiency testing is a form of external quality control for an assay. It is usually administered by a reference laboratory that distributes panels of samples, receives the results from the laboratories, analyses the data, and reports the results back to the laboratories. If results from an assay at a given laboratory remain within acceptable limits and show evidence of accuracy and reproducibility, the laboratory may be certified by government agencies or reference laboratories as an official laboratory for that assay. Panels of sera for proficiency testing should contain a full representation of an analyte's concentration in animals of the target population. If the panels only have high-positive and low-positive sera (with none near the assay's cut-off), the exercise will only give evidence of reproducibility at the extremes of analyte concentration, and will not clarify whether routine test results on the target population property classify infection status of animals.

Because of the extraordinary set of variables that impact on the performance of serodiagnostic assays, it is highly desirability to expand the number of standard sera from animals of known infection status because of the principle that error in the estimates of D-SN and D-SP is reduced with increasing sample size. Furthermore, when the assay is to be transferred to a completely different geographic region, it is essential to re-validate the assay by subjecting it to sera from populations of animals that reside under local conditions.

When serum control samples are nearing depletion, it is essential to prepare and repeatedly test the replacement samples. When other reagents, such as antigen for capture of antibody, must be replaced, they should be produced using the same criteria as for the original reagents, and tested in at least five runs of the assay. Whenever possible, it is important to change only one reagent at a time to avoid the compound problem of evaluating more than one variable at a time

## C. VALIDATION OF ASSAYS OTHER THAN ENZYME-LINKED IMMUNOSORBENT ASSAY

Although the example used has been an indirect ELISA test, the same principles apply to the validation of any diagnostic assay. It is of utmost importance not to stop after the first two stages of assay validation - that does not constitute a validated assay for diagnostic use. Although reagent and protocol refinement are important, the selection of the reference populations is probably the most critical factor. it is no surprise when reviewing the literature to find a wide range of estimates for D-

SN and D-SP for the same basic assay. Although part of the variation may be attributed to the reagents chosen, it is likely that the variation in estimates of D-SN and D-SP is due to biased selection of sera on which the test was 'validated'. This stage in assay validation needs more attention than it has been given previously. This is particularly true in the current atmosphere of international trade agreements and all their implications with regard to movement of animals and animal products.

## REFERENCES


[1] WRIGHT, P.F., NILSSON, E., VAN ROOIJ, E.M.A., LELENTA, M., JEGGO, M.H., Standardisation and validation of enzyme-linked immunosorbent assay techniques for the detection of antibody in infectious disease diagnosis, Rev. sci. tech. Off. int Epiz. **12** (1993) 435-450.

[2] CROWTHER, J.R., ELISA theory and practice. In: Methods in Molecular Biology. Humana Press, Totowa, NJ, USA, (1995) 1-256.

[3] JACOBSON, R.H., Validation of serological assays for diagnosis of infectious diseases, Manual of Standards for Diagnostic Tests and Vaccines, OIE, France (1996) 8-15.

[4] GREINER, M., SOHR D., GOBEL P., A modified ROC analysis for the selection of cut-off values and the definition of intermediate results of serodiagnostic tests, J. Immunol Methods **185** (1995) 123-132.

[5] ZWEIG, M.H., CAMPBELL, G., Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine, Clin. Chem. **39** (1993) 561-577.

[6] SMITH, R.D., Clinical Veterinary Epidemiology, Butterworth-Heinemann, Stoneham, MA, USA, (1991) 1-223.

[7] GREINER, M., FRANAKE, C.R., BOHNING, D., SCHLATTMANN, P., Construction of an intrinsic cut-off value for the sero-epidemiological study of *Trypanosoma evansi* infections in a canine population in Brazil: a new approach towards unbiased estimation of prevalence, Acta Trop. **56** (1994) 97-109.

[8] JACOBSON, R.H., How well do serodiagnostic tests predict the infection of disease status of cats, J. Am. Vet. Med. Assoc. **199** (1991) 1343-1347.

[9] CEMBROWSKI, G.S., SULLIVAN, A.M., Quality control and statistics, In: An Introduction to Clinical Chemistry, Bishop, M.L., Duben-Engelkirk, J.L., Fody, E.P., eds. Lippincott, Philadelphia, USA, (1992) 63-101.